

---

# CONTEMPORARY ISSUES IN FORENSIC HANDWRITING EXAMINATION. A DISCUSSION OF KEY ISSUES IN THE WAKE OF THE STARZECPYZEL DECISION

Bryan Found<sup>1</sup>, Doug Rogers<sup>1</sup>

---

**Abstract:** *A considerable amount of attention has been focused on the field of forensic handwriting examination as a result of a recent Daubert hearing regarding the admissibility of forensic handwriting testimony (United States of America v. Roberta and Eileen Starzecpyzel, 1995). The findings of the hearing provide us with an opportunity to reflect on some of the basic postulates and practices associated with the field, particularly as they are perceived by individuals working within mainstream scientific paradigms. It appears that there are some postulates that are still mounted as underpinning forensic handwriting examination that defy even basic logic when seen in the environment of normal behavioural sciences. Rather than dwell on the possible reasons for this phenomenon, a few basic alternatives to the current explanation of theory and practice will be overviewed. Although what is presented here is largely 'theoretical' in nature, it does provide a framework which currently forms the focus of our research. Ultimately, the question as to whether what we do can be regarded as science or a practical skill falls within the frame of reference of those who choose to define those terms. What is important is not that we waste time and effort arguing over the details of which group we belong to, but rather that we concentrate on improving the paradigm within which we all work. The first step in this process is defining what the paradigm is.*

---

**Reference:** Bryan Found, Doug Rogers (1995, Vol. 8 – reformatted and reprinted). Contemporary issues in forensic handwriting examination. A discussion of Key Issues in the Wake of the Starzecpyzel Decision. J. Forensic Document Examination, Vol. 29, pp. 5 - 22.

**Keywords:** Daubert Hearing, feature detection, expressing opinions, complexity theory, similarities, differences

---

## 1. Introduction

“The Daubert hearing established that forensic document examination, which clothes itself in the trappings of science, does not rest on carefully articulated postulates, does not employ rigorous methodology, and has not convincingly documented the accuracy of its determinations” (US v. Starzecpyzel, 880 F. Supp. 1027, [SDNY.1995]). This statement highlights major problems associated with the field of forensic handwriting examination.

It includes criticisms that have been made in other articles (Huber & Headrick, 1990; Risinger, Denbeaux & Saks, 1989) concerning the science of forensic handwriting examination and associated issues of method and validation. There is yet to be a standard text from which we have been able to extract clear statements of what can reasonably be said about handwriting, together with an accompanying theoretical basis and a study of validation. Nevertheless, in the ‘Memorandum and Order’ Judge McKenna stated that “Saks’ testimony established that there is no strong statistical evidence supporting or disproving the ‘two fundamental principles’ or the reliability of forensic document examination”. There is not, therefore, a suggestion that the practices of

---

1.National Forensic Handwriting Consultancy,  
Handwriting Analysis and Research Laboratory,  
School of Human Biosciences, La Trobe University,  
Locked Bag 12, Carlton South, Victoria, 3053,  
Australia

the field can't be done, but that there is a lack of an accepted theoretical basis on which we conduct our work and an absence of proof of our reliability. If we are to be recognized as adhering to the process of science, this theoretical basis must be supported by appropriately designed research, and the application of the resulting theory must then be validated. In the scientific environment, validation studies do not refer to case examples or even the features associated with known forgeries, for example, but rather to extensive and realistic tests of examiners to produce the correct result when the true answer is not known to them. There is no question that there has been a significant lack of these classically-designed validation trials. As a profession we are responsible for this shortfall and should heed the criticism, regardless of its source, in a professional manner.

Primarily it appears that what is currently most inappropriate is the image of what is done in the field under the banner of science. This is reflected in the statement that, "The problem arises from the likely perception by jurors that FDEs are scientists, which would suggest far greater precision and reliability than was established by the Daubert hearing. This perception might arise from several sources, such as the appearance of the words *scientific* and *laboratory* in much of the relevant literature, and the overly precise manner in which FDEs describe their level of confidence in their opinions as to whether questioned writings are genuine." Unfortunately, there is an underlying assumption that all document examiners conduct the work on the same basis that was suggested in this hearing. We do not, and certainly do not suggest that all others do.

The creation of a science of handwriting analysis as was suggested by Judge McKenna is, although young in forensic terms, already having an impact. An example of this type of approach comes from the joint conference of the International Graphonomics Society and the Association of Forensic Document Examiners held in Canada in 1995. In addition, measurement techniques and criteria developments specific to forensic handwriting examination have been reported on (Cheung & Leung, 1989; Baier, 1995; Found & Rogers, 1995; Found, Rogers & Schmittat, 1994; Found, Rogers, Metz & Schmittat, 1994). Part of our treatment of handwriting

examinations has been to attempt to standardize and document handwriting methodology (Found & Dick, 1992; Found, Dick & Rogers, 1994; Metz, Found, Dick & Rogers, 1995). The primary change to existing technique has been reporting procedures, which necessarily have been made to reflect both the considerable limitations associated with the type of material being examined, and the need for clarity of meaning of opinion in the court environment. This process is, of course, very slow due to the normal resistance to change, lack of research time and money and suitably qualified individuals devoted to the field. As was noted, "...this discipline has no counterpart in industry or academia with an economic incentive to study and refine its scientific basis"). The handwriting examination component of document examination has largely drifted and not developed at the rate that normal science would have expected. The field of forensic handwriting examination, for these and other essentially theoretical reasons, falls well short of the *identification science* that it has commonly been perceived to be. Indeed, evidence based on the outcomes of human movement cannot and should not in any way be paralleled to forensic fields such as DNA and fingerprints. It could be argued that the severity of criticism that we have been subjected to is probably related to the power that this branch of forensic science has claimed. The claim is simply not supported in theory, nor have we supplied the evidence in practice. We as a group are responsible for this reality. We are, however, like those before us, only transient in this process. We have a choice to either participate in reconstructing and validating the discipline such that its value, if we find it to have value, is maintained for those who follow us.

This paper aims to deal, in a general way, with some of the issues raised in the Daubert hearing which impinge upon the above major concerns or criticisms. To review in any exhaustive fashion what was said in that hearing, as well as the conclusion of the court, would be too extensive a task to explore here in any meaningful way. Indeed, we found it an impossible task as handwriting specialists, given that much of the questioning was based on statements of underlying beliefs and reporting formats that, although they appear to have gained general acceptance in the forensic community, we do not agree with.

## **2. Identifying key issues**

There were a number of issues brought out during the hearing that we feel should be developed. These are the notion of individual and class characteristics, issues regarding similarities and differences and reporting procedures. Our aim is not to analyze the question and answer process associated with the hearing, but rather to discuss in general the criticisms that were raised in the context of the rationale for the methodology of the forensic comparison conducted in our laboratory.

## **3. Class and individual characteristics as a basis for handwriting opinion**

The belief in the notion of class and individual characteristics has remained a pillar in forensic handwriting examination and appears to be used as the fundamental basis by which handwriting examiners claim they can identify an individual (Conway, 1959; Harrison, 1958; Hilton, 1982; Osborn, 1929). The following passage, extracted from a prosecution's handwriting expert, also indicates that it underpins the evidence she was giving: "...you have a familiarity with the copy book standards that are being taught and you can evaluate the letter forms on how much they diverge from the standard to get an idea of how unique that is. You also understand the uniqueness of different letter forms or a particular quality of a writing based on the study you have done of the literature and of the treatises and once again drawing on your own experience in previous cases that you have also examined." The fundamentals of the class/individual theory are represented here by the notions of copybook form, divergence from the form and the assessment of uniqueness of characteristics based on experience. We do not intend to exhaustively restate the theory here, as it can be found in various forms in most of the standard texts in the field. However, basically it is claimed that the validity of a document examiner's opinion is based on his or her ability to distinguish between what are class and what are individual characteristics. There is some sort of assessment of the uniqueness of the features based on an individual's knowledge of character manifestations and combinations in the population. It has been argued, however, that although it appears to make

sense superficially, there is limited theoretical and/or practical support for it (Lacey and Dick, 1992). Some of the problems with the theory as we perceive it are outlined below.

1. No evidence has been provided that experience from doing forensic casework increases the examiners ability to differentiate between class and individual characteristics.
2. No evidence has been provided that experience increases the validity of findings.
3. Even given this theory, some handwriting specialists believe it is possible to examine and express opinions as to the authorship of foreign writings.
4. Given this theory for handwriting, signatures are somehow included, even though they may exhibit no class characteristics whatsoever and the uniqueness of the features in the image have no way of being assessed according to the theory.

So how is it that such a theory has survived? It appears to make sense when explained to the layperson and it provides a platform on which expertise can be claimed and on which one's position within the field can be improved. In addition, it is not directly falsifiable, as we have no database on which an individual's judgment of uniqueness can be validated. The theory can, however, be indirectly tested. The simple test for any person claiming to have the knowledge base to construct an analysis on the basis of this theory is as follows: 1.) Select two equally experienced examiners from a forensic laboratory and provide them with the identical sample of handwriting of a number of individuals where the class system is known. 2.) Ask them to individually determine each of the class systems and then list and rank the individual characteristics according to their degree of uniqueness in the population. The results could then be compared and correlated. We think that the results would not justify the apparent enthusiasm for the theory. This type of validation trial has been discussed with numerous document examiners and yet there has been no race to conduct the experiment.

There is, however, a place in the profession for class/individual theory. We apply the theory for the purpose of explaining to the lay person the process by which inter-and intra-writer variation emerges. There can be no reasonable grounds on which to doubt that handwriting is normally learnt in the first instance by reproducing a copy book system. It is common knowledge that individuals introduce into their writing, either consciously or subconsciously, additional features or modifications on that copy book form for a whole range of reasons such as increased speed production, incorrect adherence to the system, changing the aesthetic, qualities etc. The problem arises as a result of the belief by some handwriting experts that they can retrospectively determine the source of the components of the graphemes and then claim that divergent characteristics can be subjectively weighted as to their respective uniqueness and individualising power. Since the theory is not supported logically, is able to be tested but has not been, and the basis of opinion relies on information that is individual specific and not falsifiable, it does not sit easily within a scientific paradigm.

We suggest that we can do no less than either modify or abandon this theory. But is there an alternative theory which can be validated and on which opinions can be mounted that makes sense? The reality is that there may be a variety of theories that could be proposed. We choose to rationalize the examination process and the underlying logic, not according to the determination of significant individualizing characteristics, but rather to the determination of overall similarity or difference associated with observable features and basic relationships which are thought to exist between the underlying physiological mechanism responsible for the image, the variation that is observed in image production in the population, and the observed difficulty that individuals have in copying complex movements.

#### 4. Modifying the theory

We can propose a basic model of the forensic comparison method that we conduct in our laboratory, a simplified version of which is represented in Figure 1. Fundamentally it is a comparison where the resultant first stage of the opinion, similar to traditional approaches, deals with the notion of

similarity or difference. It is fairly straightforward to advance a plausible explanation once we have made a decision about this if one is able to be made. The decision arrived at should be understandable, logical and illustratable to any impartial person. The legal system rightly tends to focus on this stage of the examination because of its subjectivity and the resultant implications to the conclusions regarding the dispute. The notion of significant similarity or difference will be elaborated on later. There are a number of aspects of this particular process that we feel should be discussed.

One of the most difficult aspects when reflecting on visual comparison processes is to explain exactly how it is that our brains are processing the information that we are providing it with. There is great difficulty in verbally describing what our brains judge to be similar or different. Since we are dealing with a visual phenomenon, sense can only be made of the concept according to visual illustration. Semantic gymnastics on this point, of the type observable in the Daubert hearing, result directly from this phenomenon. There did appear to be some confusion at this point regarding distinguishing inter-writer differences from natural variation. However, in terms of the approach outlined here, this is not the stage where that distinction is considered. Decisions as to overall similarity or difference are about all of the elements of the image, from line details, character constructions, character combination constructions, word constructions and features associated with the entire text. No significance as such is attached to this opinion. Judge McKenna did not dispute the ability of document examiners to express an opinion regarding this stage of the examination: "Although Ms. Kelly was unable to explain to the Court's satisfaction precisely how significant similarities or differences were identified, the Court has no doubt that such identifications can be performed, in some cases by cursory examination." Attempting to verbally describe this process is analogous to describing the difference and similarities between two paintings of an identical scene, but where specific paintings have not been provided to the audience.

Confusing the issue of image comparison is the usual tendency of both document examiners and the legal fraternity to focus discussions on the process in terms of character formations. We talk about g

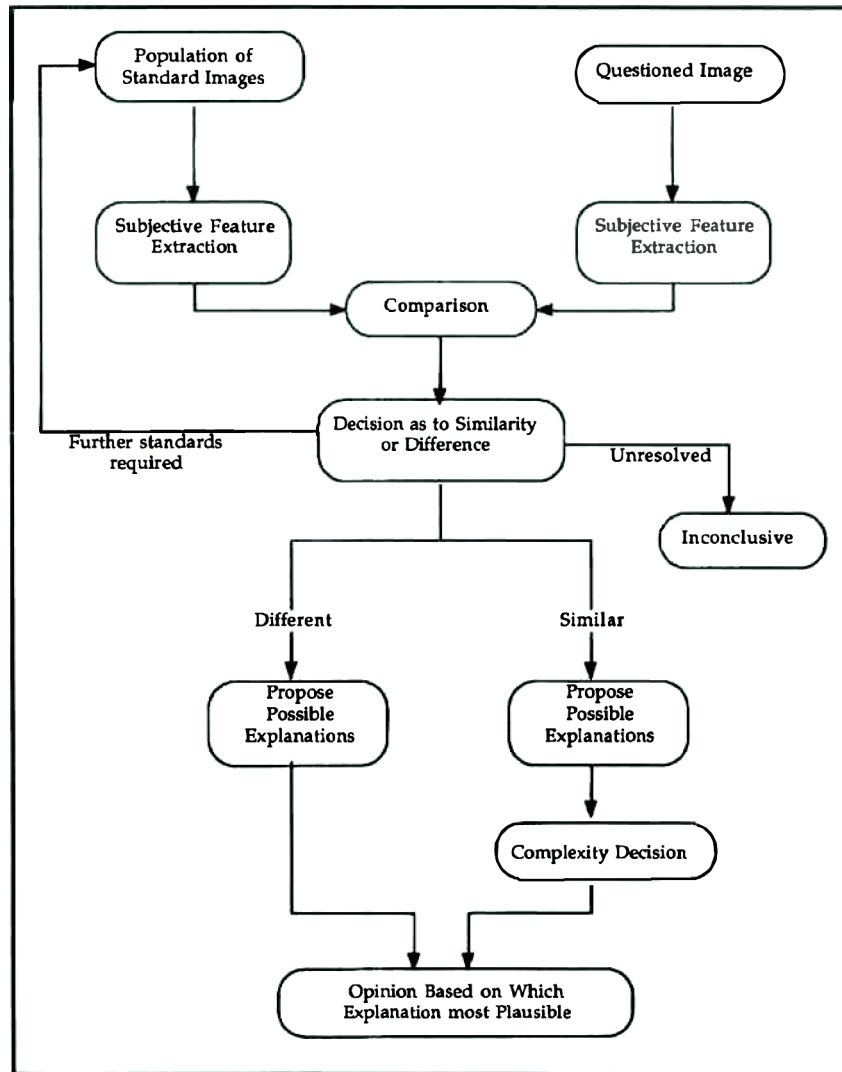


Figure 1. Flow diagram of the stages in the forensic examination of handwriting.

formations, *e* formations, proportions of staffs on different characters, etc. Writing is thought about in this way because of the relationship between the static image and its purpose, which in most instances is for the writing to be read.

It is most convenient, of course, to structure the analysis process according to these visual/linguistic cues. However, there is the hazard that this emphasis could be misconstrued to mean that these characters form the fundamental basis of the examination and opinion process. Characters can be considered the middle ground of the overall comparison and provide us with a reference point to make the comparison process manageable, particularly when we have extended text. Clearly, we are making inferences at the first stage of the examination process as to whether the

images, the artifacts of the human movement system, are the product of similar or dissimilar movement commands. The characters themselves are simply fabrications of the movement system, given significance only in light of their value in communication. This is, of course, not what forensic handwriting examination is about. We are attempting to determine whether any meaningful statements can be made purely on properties associated with the movement outcome itself. The purpose of mentioning this at this stage in the discussion is that the misunderstandings associated with this concept continue to support the enthusiasm for the construction of handwriting characteristic databases. Databases, when appropriately constructed and used, can be very powerful, particularly in systems where the construct characteristics of the file are easy

to isolate, where the substance being filed has relatively invariant properties, and where the population being sampled is relatively static and invariant itself (such that the database can be considered a reasonably representative sample). DNA and fingerprints, to differing extents, suit a paradigm revolving around significance determined from databases. Handwriting, however, does not. The nature of handwriting as a changeable outcome of learnt human movement violates each of the requirements for databases if we choose to use them to make statements concerned with statistical individualizing power. Databases such as the FISH system, which have been in use for many years as recently described by Baier (1995) and Hecker (1995), have not been reported to be used for this purpose and should not be erroneously included in this debate. We can only guess at the court's response to quoting frequencies of characteristics and significance in view of these limitations.

### 5. Feature detection

Our method is underpinned by an approach to the examination of handwriting which we have coined *feature detection*. Feature detection is based on the rationale that, under normal conditions, given a sufficient amount of writing, no two skilled writers are likely to produce handwritten images that are exactly the same in terms of the combination of construction, line quality, formation variation and text structure features. This statement is different from that offered as one of the two basic principles in the Daubert hearing that, "no two people write exactly the same way." The underlying principle associated with this theory is quite appropriately heavily qualified, and the limitations which impose this qualification should be expressed along with any findings. Basically, however, we would argue that if we were to select at random any number of extended handwriting samples from the general population, the incidence of samples that share exactly all combinations of features should be low. There is evidence for this, although criticisms regarding this notion not having been proved in a scientific way are quite valid. The basis of the working hypothesis of inter-writer difference comes from a variety of sources:

1. That handwriting is a learned behaviour involving very complex manipulations of muscles by the nervous system. As with any skilled movements, people are observed carrying them out in different ways to achieve what are often very similar goals; e.g., playing a sport, talking, playing a musical instrument, painting, etc. The reality is that it is accepted that the outcome of these movements differs from person to person and in the skilled 'mover' may result in a movement style that is to some extent characteristic. It is no different with handwriting. We commonly see evidence of this through the course of our lives, through recognizing the writing of our wife or husband or workmate. The problem is that we do not have significant numerical support for this notion.
2. There has been no report of extended writings that are exactly the same, even though the field of forensic handwriting examination has been operating in an organized way for decades, and databases of handwriting samples are kept in some form at many government laboratories. Databases associated with anonymous letter files are routinely searched on some basis. Those who have had to carry out this process indicate that it is not difficult because of the vast array of ways that writing presents itself.
3. If the handwriting of individuals was commonly similar and the pictorial results of the movements were easy to reproduce, the commercial world should have experienced anarchy by now as a result of the ease with which funds could be fraudulently withdrawn.
4. Instruments such as the FISH system would be of no value, as the search strategy which relies on healthy inter-writer variation would invariably throw back at the operator an unmanageable sample of potential hits. Yet the FISH

system has survived and research using the databases is still being reported (Baier, 1995; Hecker, 1995). Other handwriting classification schemes have also been developed (Hardcastle & Kemmenoe, 1990; Hardcastle, Thornton & Totty, 1986), the latter of which contains further references to these schemes.

5. We would not see the level of research being carried out on optical character recognition devices. Examples of this type of research appears in the Proceedings of the Third International Conference on Document Analysis and Recognition, 1995. One of the major problems for these devices is the vast number of ways that handwriting presents itself, both inter- and intra-writer.
6. There have been a number of studies carried out that, although they focus on only a small quantity of writing and only a limited number of characteristics, still provide evidence of this variation (Eldridge, Nimmo-Smith & Wing, 1985; Livingston, 1963; Muehlberger, Newman, Regent & Wichmann, 1977; Franks, Davis, Totty, Hardcastle & Grove, 1985; Wing & Nimmo-Smith, 1987).
7. Perhaps one line of support for this notion of inter-writer variability is the inability to support the alternative hypothesis that most people write the same as each other. One can only wonder what a court's response would be if we stood up and claimed that most people write exactly the same as each other. This flies in the face of common knowledge.
8. There have been reports of large-scale handwriting searches that have been successful in isolating individuals, even on the basis of limited comparison strategies (Baxendale & Renshaw, 1979; Harvey & Mitchell, 1973). Although there are only a small number of published reports, strategies of this type are not uncommonly put to use.

It is not unreasonable to accept inter-writer variation as a working hypothesis, even though it has not been delineated mathematically. In addition, the second principle stated in the Daubert hearing that, "no two people will write exactly the same when repeating," although it should not have been stated in such absolute terms, is able to be observed and reasonably explained. It results from a combination of an individual's motor output varying to different extents due to the non-muscle specific nature of the movement's representation in the brain (Van Galen, 1980), personal tolerances of motor output, the relative position of the movement system when the entry is to be executed, changes associated with particular character combinations, or conscious changes to the movement process.

The inter-and intra-writer variation can be thought of as a product of these factors. Given this breakdown, it is not surprising that persons in the general population recognize easily familiar writings and routinely conduct their own handwriting examinations. We could argue that it is elements of the *picture* of the writing that their brains are comparing to a given number of known *writing pictures* stored in memory. These writing pictures are laid down by constant exposure to the handwriting of others. For this recognition to be achieved, the brain must be making a decision based on *pictorial features*, or more probably a range of them, within the writing. In this situation the brain may be excluding alternate pictorial memories where the features do not match, in favour of those that do. It is plausible, therefore, that the brain is making decisions based on those features that pictorially characterise the writing.

This process is relatively straightforward for a member of the general population, as only a limited number of pictorial memories are referred to and an incorrect judgment may have no implications. The writing is then either judged as *known* or *unknown*. Handwriting examiners are faced with a different situation in that every sample of writing submitted is unknown. Collected or requested handwriting standards are then used to form a working knowledge of the writer suspected of writing the questioned entries. The gathering of handwriting standards is covered adequately in the texts and will not be further discussed here. The question is, 'On what basis is the

handwriting being compared and what is the nature of the expertise that is claimed?"

Feature detection rationalizes that, given an adequate quantity of skilled standard and questioned writing, the brain can perform an analysis of the standard writing and determine either visually or using magnification, spatial features or line quality features which contribute to the writing's pictorial character. It is these features that are being compared to the questioned writing. It is on the basis of these features that the primary opinion is formed as to whether the body of questioned material is similar to or different from the body of standard material. There is no speculating as to whether characteristics are class, individual or a combination of both.

### 6. Similarity or difference

Perhaps one of the most confusing of concepts in our field is the explanation of what in writing constitutes a similarity or a difference, particularly in light of the variation phenomena. In terms of our model, we define the terms generally; similarities are pictorial or structural features that appear consistent between the populations of questioned and standard images. The similarities can be observed in terms of the way the strokes are concatenated into letter, letter combinations and word formations, the features that can be described, and the relative placement of images. Differences are pictorial or structural features that appear dissimilar between the populations of questioned and standard images. The dissimilarities can be observed in terms of one, or combinations of the way the strokes are concatenated into letter, letter combinations, word formations and the features that can be described. The criteria for features to be described as different is that they are fundamental to the pictorial or structural character of the writing and are not shared between the bodies of questioned and standard writings. Examples of differences would be a character which is consistently constructed in a different way between the questioned and standard images, or where the line quality is visually dissimilar between the questioned and standard images etc.

Clearly, these definitions do not address issues of authorship. What they do, however, is to focus the examination on the appropriate set of hypotheses. What is important is that a decision at this stage in

the methodology is illustratable. In many instances, the comparison process stalls at this point and a reasonable opinion cannot be formed as to difference or similarity. This results in an inconclusive result.

At this stage there is simply no numerical answer as to what is an adequate amount of known or questioned handwriting. This remains another limitation of the examination technique which must be respected. We can, however, show in specific examples why in our opinion there is an insufficient amount and why in another example there is sufficient.

### 7. Expressing opinions based on observations of similarity or difference

Given that we have subjectively formed an opinion as to whether the questioned material is similar to or different from the standard material, we can now propose explanations that could account for that primary observation. The ultimate aim is to express an opinion as to which of the alternative explanations is the most plausible. This process should always be carried out in an environment where no other peripheral information is taken into account. Peripheral information belongs to the investigators and to the courts. We can certainly be asked in the courtroom how certain factors may effect handwriting, but this should not contaminate our perception of what we can reasonably accomplish dealing solely with the handwritten images. Let us consider a typical example.

Imagine that we have performed an analysis of a questioned signature. The opinion of the examiner is that there are no differences in the line quality, construction or spatial characteristics when compared to the population of standard material. We could conclude from this that the image is the product of the same or similar movement commands or different movement commands that produced what would appear to be an artifact consistent with the population of standard signatures. These statements are not about authorship. We can develop on these statements to propose three explanations or hypotheses to explain these similarities in terms of authorship. This section of the method was referred to by Judge McKenna as "...the second stage of their analysis where FDEs combine their first stage results and draw inferences as to the genuineness of questioned signatures".



The three explanations that we propose are:

1. The questioned signature was written by the writer of the standard material.
2. The questioned signature was simulated by a writer other than the standard writer such that no evidence of the simulation process remains.
3. We have a chance match between the questioned signature and another person's signature.

We could, if we choose, stop at that point and let the court make a ruling as to which of these explanations is acceptable beyond reasonable doubt or in the balance of probabilities. Of course the court in many instances will have a great advantage over the document examiner, as other evidence can be brought in which may change the plausibility of any one of these explanations. We would argue, however, that the expertise of the document examiner can still be applied at this stage. The expertise required to do this, however, is not based on properties such as the determination of uniqueness or individual characteristics, but rather is derived from a number of fundamental relationships that we propose exist and beg further investigation. What follows is an explanation of these relationships.

### 8. Complexity theory

Skilled handwriting is thought to be manufactured by a series of concatenated single curvilinear strokes. The function of the motor system, summarized by Thomassen and van Galen (1992), although relevant to the underlying theory, will not be detailed here. What is important is that in skilled writers, underlying kinematic order is observed amongst individuals. In the absence of this order we would be unable to carry out any sort of examination based on theories such as are being proposed. Obviously there is a relationship between characters, the concatenation of strokes and the underlying physiological system. When handwriting examiners draw out features, what they are doing is describing the relationship between the participating strokes in the resultant character which may describe the shape or construction of a complete character, sections within it, or relationships

between them. There theoretically is, with a skilled writer, a relationship between the number of these stroke concatenations, the resultant features, and the complexity of a given sample of handwriting. It is the notion of *complexity* that is central to our method, enabling opinions to be expressed regarding authorship.

Complexity of handwriting can theoretically be related either singularly or jointly to a whole range of characteristics resulting from differing orientations of concatenating strokes. Examples of these resultant characteristics may be the total length of the line, the number of points where the line exhibits feathering, the degree that the line is superimposed on itself etc. We propose that there are a number of basic relationships that exist which enable opinions to be expressed about any nexus that may exist between questioned and standard writings, once the decision that the questioned writing is similar to or consistent with the standard writing has been made. These theoretical relationships can be investigated using normal scientific validation protocols. These relationships are described in figures 2 to 4. For clarity, the general logic underlying these relationships will be described.

#### 8.1 The number of concatenated strokes versus the complexity

The first relationship is the number and relative orientation of concatenating strokes, or a measure of this parameter such as the number of curvature maxima, as a predictor of complexity. That is, in the skilled writer, the greater number of times the pen was required to change direction without a penlift, the more visually complex the image appears.

#### 8.2 The complexity versus the likelihood of a chance match

This relationship follows from that stated above in that, given that all writings share common components such as concatenating strokes, and given that the number of concatenating strokes contribute to the complexity, then if we were to choose random samples exhibiting identical text, as we proceed through an analysis of the concatenations, the complexity increases and so does the likelihood that the samples will diverge in some way from each other.

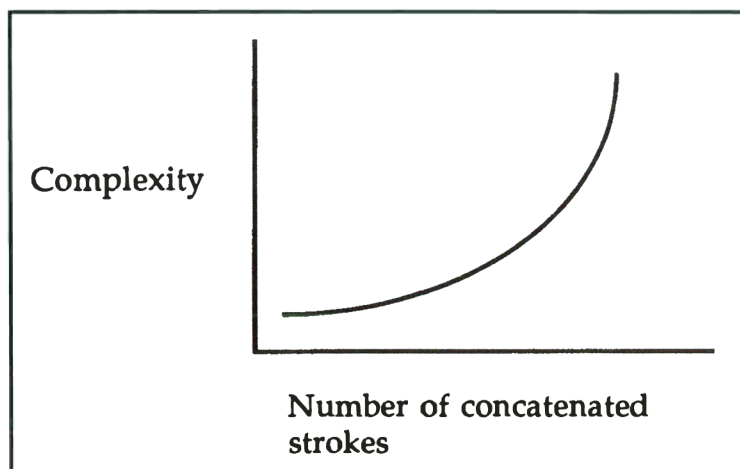


Figure 2. The theoretical relationship proposed between the number of concatenated strokes (or a measure thereof) and a handwritten image's complexity.

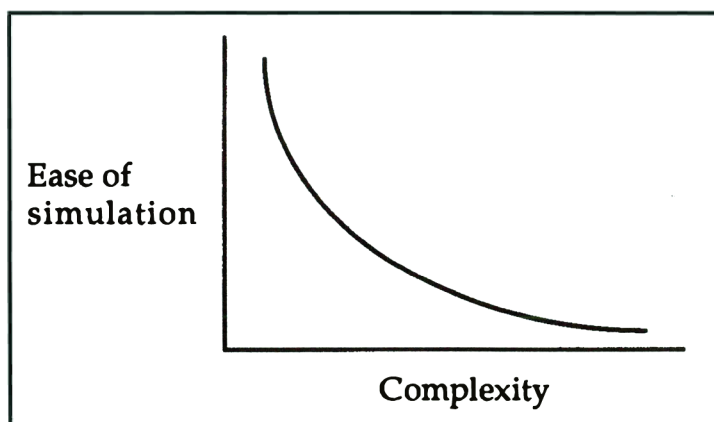


Figure 3. The theoretical relationship proposed between the complexity of a handwritten image and the ease with which it could be simulated successfully.

### 8.3 The complexity versus the ease of simulation

Given the above, as the image becomes more complex, it would make sense that it would be more difficult to simulate. An example of this would be copying a straight line in comparison to copying an extended section of text.

The issue becomes not one associated with the frequency of feature formations in the population or the subjective assessment thereof, but rather, if we accept that most individuals write differently from one another, the complexity of the static image. The research direction is therefore investigating the questions: What evidence do we have which supports the proposed relationships? What features best predict

a written image's complexity? How can we objectively measure complexity predictors? There are a number of ways to investigate this phenomena. One way is to get handwriting experts to group images according to their perceptions of complexity and then to analyze the image according to characteristics that can be counted or measured objectively. This approach has been reported on, but not for the reasons as stated here (Found & Rogers, 1995). Another method is to correlate parameters measured for handwritten formations with a measure of success in actually forging these characteristics. The latter has not as yet been attempted by the authors but is feasible if undertaken in an objective manner.

The complexity theory also enables us to explain the common ground between handwriting text base

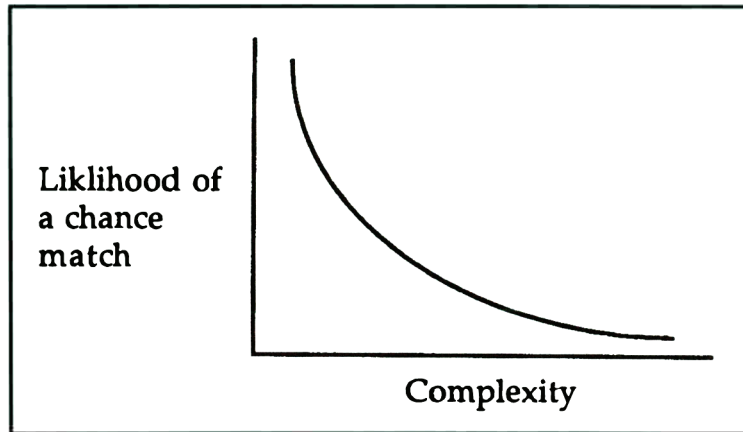


Figure 4. The theoretical relationship proposed between the complexity of a handwritten image and the likelihood of a chance match between the handwritten features of any two individuals

examinations and signature examinations because we have diminished the importance of the *linguistic* cues. The issue of whether we should attempt to examine foreign writings, however, remains questionable because of the difficulty in constructing the general sense of the examination through letter, letter combinations and word cues. This issue is associated with method and will not be further discussed here. Perhaps this problem will only be overcome when the human factor is completely removed from the examination equation.

Complexity, we believe, may also be the key to defining how much text is needed to express a valid opinion. Objective means to quantify the best predictors of complexity may also provide us with better definitions of what constitutes similarities or differences.

Given the underlying order of handwriting production in terms of concatenating strokes, the proposed relationship between complexity, and the likelihood of chance match and ability to be simulated, we can now address the explanations given for the similarity of the questioned and standard signature given in our example in terms of the theory. An example of the rationale for expressing an opinion that there exists a nexus between the questioned and standard writing is as follows:

1. Explanation number 3 is in our opinion implausible on the basis that it is not believable that a signature falling within the range of variation in the standard

material has been incorporated into the document by accident, or that an individual attempting to mark the document fraudulently has by chance produced a signature appearing to be genuine, even though that was not the intention. We do not, therefore, support this explanation.

2. Explanation number 2 is in our opinion implausible on the basis that the signature appears to be complex, fluently written and not bearing any indicators of a simulation process. The opinion of the examiner is that the signature exhibits sufficient features that would bestow upon it a measure of difficulty if it were attempted to be simulated. We do not, therefore, support this explanation.
3. Explanation number 1 is, therefore, the only remaining explanation. Given the complexity of the image and the absence of differences, it is considered the most plausible hypothesis. We therefore support this hypothesis.

Therefore, it is the opinion of the examiner that the most plausible explanation accounting for the similarities observed is that the writer of the standard signatures also wrote the questioned signature.

In this way we can still express an opinion as to

the authorship of the questioned entry. However, it is made quite clear that there are alternatives that have to be recognized. These alternatives can never be absolutely excluded due to a combination of factors including the nature of behavioural artifacts, the lack of objective techniques available to analyze them, and the inability to meaningfully support or exclude them statistically. We do not express the results in probabilistic terms, but only on our beliefs according to this method and expertise in applying it. Of course there will be many instances where there is not a clearly most plausible hypothesis. These cases are inconclusive. It may be possible to suggest limited support for one explanation over the others on the grounds that, for example, there is limited standard material such that all of the features could not be accounted for.

The process outlined above is relatively straightforward when dealing with a case where the questioned entry is similar and complex. There is a tendency to believe that these sorts of processes should work equally well in both directions. They simply do not. This illogical belief is reflected in some reporting procedures where opinions are stated to range from a total identification of a given person to a questioned document, to the total elimination of that person as having authored something. This can pose some difficulty in court, particularly when it is commonly touted and believed that forensic handwriting examination is equally as good at excluding individuals from having written a questioned document. Let us proceed through the same hypothetical situation as before. Let us imagine that we have performed an analysis of questioned handwriting. The opinion of the examiner is that there are differences in the line quality, construction and spatial characteristics associated with the questioned handwriting when compared to the population of standard material. Given this situation, there exists only one general explanation that could be advanced to explain the differences: the questioned signature was unlikely to have been produced using the same neuromusculature commands as were used to form the standard writing.

As can be observed, this is not a statement about beliefs as to authorship. We now must look at possibilities that could account for this primary finding. Examples of these explanations are:

1. The standard writer did not write the entries.
2. The population of writings submitted as standard is not representative of the standard writer's normal handwriting and the standard writer was responsible for the entries.
3. The standard writer is capable of producing more than one writing style.
4. The standard writer has purposefully changed his or her writing.
5. The writing of the standard writer has been affected by unknown internal or environmental factors. Examples of these factors are age, illness and intoxication, references to which can be found in Ellen 1989( p.45).

The greatest problem that we have in this situation is providing support for one of these plausible explanations over all of the others. It is very difficult to justify the opinion that the standard writer did not write the entries, as to do this one must be able to illustrate that that writer was incapable of having written the entries. In addition, we must also provide meaningful research that would justify not supporting the alternatives. Although through research (e.g. effects of alcohol on writing) we may be able to state general trends, there are real threats to external validity on applying these results to any specific case example. With a great number of standards and questioned material taken from around the same time, it may be possible to reasonably provide limited support for hypothesis 1, but given the nature of the alternatives, exclusion would be inappropriate.

This should not, of course, be seen to detract from the evidential power of handwriting examinations. The opinion that the standards and questioned entries are different in a major and illustratable way does provide the court with information that may be of use given the other lines of evidence.

The relationship between handwriting analysis and exclusionary opinions becomes more distant as we move down the scale of complexity as illustrated in figures 3 and 4. At the lower end of this scale we

have not only spurious signatures and small amounts of questioned writing, but also larger samples of writing that are not considered to be skilled; that is, writings with poor line quality or writings that are simply or variably constructed. If we look at the same plausible alternatives to explain differences with signature formations, then we are faced with the situation that it is almost impossible to support any one of the alternatives in a meaningful way. The misunderstanding of this concept is seemingly illustrated in an article by Beck (1995) who stated that, “The principle of elimination is as simple as basic scientific method: no matter how much evidence exists for a theory, it must be rejected if even a single significant contradiction is encountered.” This discussion then proceeds to support this statement on the basis of other statements by Harrison (1958), Osborn (1929) and Hilton (1982). In this case, we consider the logic is being applied to the wrong section of the methodology. The *theory* in this instance is at the level of whether the questioned material is similar to the standard material. If a ‘single significant contradiction’ is encountered, then we would agree that the opinion would be that the bodies of writing are different, not that the writers of the bodies of writing are different as is stated.

### 9. Reporting procedures

Having established the subjective nature of the examination process and the limitations imposed by the underlying theoretical framework, we now must consider how best to express the results of our analysis. The interface between what we do and what the perception is of what we do and mean is conveyed primarily at this stage. Reporting procedure is diverse in the field. However, it appears that in America at least, the probability scale is popularly accepted (McAlexander, 1991). The problem with this scale is that it implies a level of exactness that is not supportable by any studies, nor by the theory underlying it. This was reflected in Judge McKenna’s comment that “No showing has been made, however, that FDEs can combine their first stage observations into such accurate conclusions as would justify a nine level scale.” In addition, there is predictable confusion between the probability terms used and the mathematics that usually underlie them in traditional scientific paradigms.

Arguably the most flawed aspect of its use on scientific grounds is the top and bottom two levels of opinion, where we have both highly probable and certainty. We would argue that, even given this system, highly probable would be the highest opinion that could reasonably be expressed because of the inability of the examiner to absolutely exclude alternative hypotheses to account for the differences and or similarities observed. Indeed, the use of the word *certainty* in the court room is most inappropriate, particularly when the subjective nature of the analysis may not have been made clear, and where the perception of the study may have been coloured by terms such as, scientific, identification, individual characteristics, experience, etc.

Fortunately, there are alternatives to expressing results according to the scales described above. The first suggestion is to make clear in written reports the limitations associated with the type of evidence that is being presented:

1. The images are the artifacts of human movement and do not in themselves define the process by which they were carried out. Indeed, the image that we examine can at best be considered a sample of the overall movement outcome. Dynamic information, although it can in some ways be inferred, is not available.
2. The written image from the same person can manifest differently, primarily as a result of the underlying neuromuscular system which is responsible for its execution. In addition, environmental factors associated with the writing implement, the writing medium, and body position may alter the artifact.
3. Handwriting, as with any learnt motor behaviour, can be modified (either consciously or subconsciously) or mimicked.
4. Although handwriting features are focused upon when making comparisons, the absolute significance of these features are not able to be determined.

The results section should contain a statement as to similarity or difference, a list of the plausible explanations that could account for this primary opinion, and a discussion as to why alternative explanations were excluded in favour of the one that the examiner is supporting.

### 10. Theory and forgery

It appears that the dispute over forensic handwriting examination in *US v. Starzecpyzel* was related to a signature case where the conclusion of a forensic document examiner was that the signatures were “not genuine.” Given that this was the starting point of the dispute, it may be appropriate to discuss the examination of static signature formations in light of the theories proposed in this paper. The example used by Judge McKenna will be used to investigate this point. These signatures are drawn from Harrison (1958). Judge McKenna states that the illustration “...shows two signatures with many identifiable differences such as the ornamentation of each “B” and the curvature of the initial stroke of each “M.” Given no other exemplars, the lay examiner might correctly conclude that one of the signatures was a forgery. While an FDE might come to the same conclusion, he or she would first have considered the possibility that both signatures were genuine, the differences arising from such sources as natural variation, the passage of time, purposeful alteration (e.g., elaborate signatures used when signing checks), illness, or intoxication. As Ms. Kelly repeatedly stated throughout her testimony, FDEs are aware that forgery detection requires an adequate quantity of genuine writings to eliminate such possibilities.

By way of clarification, let us use the definition of *forgery* or *fraudulent* signature as stated by Hilton (1982): “A forged signature. It involves the writing of a name as a signature by someone other than the person himself, without his permission, often with some degree of imitation.” This term, therefore, is not only a statement regarding non-authorship, but also intent. This approach simply does not fit within the model that has been proposed, nor the method that we use. Referring back to Harrison’s example, we can state that there are identifiable differences. We can illustrate these differences and if we chose to, could objectively measure them using specific software (Found, Rogers,

Metz & Schmittat, 1994); Found, Rogers & Schmittat, 1994; Found, Rogers & Schmittat, 1995). Having expressed the opinion of difference, we can then state that in our opinion the questioned signature was not produced using the same neuromusculature commands as was used to form the standard writing. The plausible explanations to explain this opinion can then be stated as has been discussed previously above. We are left with a set of explanations where we cannot reasonably, nor scientifically, exclude each of the alternatives in favour of only one. Fundamentally, in the example used, if we relate the questioned image to the complexity relationships, the complexity level is low and so the number of individuals that could perform the signature is high. There is no reasonable basis on which to exclude the standard writer as one of these individuals that could have performed this particular signature. This logic is mirrored in the more recent text by Ellen (1989) who states “When significant differences typical of those found when signatures or other writings are copied or discovered in a questioned signature, and are not present in any adequate number of those known to be genuine, it can be safely concluded that the signature is not the normal signature of the suspect. If it also shows a clear overall similarity to the genuine signatures, too close to have arisen by a chance match, it can be reported as a simulation, and that there is no evidence that it was made by the writer of the genuine signature. In such cases it is usually unwise to report that because it is a simulation it was not made by the person whose writing has been simulated.”

The support for the notion that *forgeries* can be identified comes from observation of known forgeries. Reports of these are found all through document examination literature. The reality is, however, that because the differences noted in the questioned signature are similar to those noted in known forgeries, it does not mean that we can instantly conclude this is a forgery and exclude the standard writer. Of course there is a body of research that indicates how it is that individuals *forge* their own signature, what happens to signatures in various states of ill health, etc. Let us not mistake this type of research as providing validation of our ability to absolutely exclude alternative plausible explanations to account for observed differences in signatures. We can and do use this research to

answer questions in court regarding general trends. However, we do not use it to exclude the standard writer from having authored the simulation, if that is the conclusion that we come to. We would argue that the most important role for the handwriting specialist in this case is to illustrate to the court that the questioned signature is different and explain what the possibilities are that could account for these differences. If asked, “Are the different features that you observed typical of a forgery process?” we can answer that they are, but that does not mean that the standard writer did not perform the entry and that there are other explanations that could be proposed. The court has the great advantage that they can have other relevant information such that they could, under certain circumstances, support the hypothesis that the signature was forged. This is, of course, the role of the court and not the document examiner.

What has then been discussed here is a limited example. Obviously there is a difference between examining the limited line trace associated with signatures and examining extended amounts of text, although the plausible explanations accounting for similarities or differences remain similar. Overall, it is the subjective nature of the entire process, coupled with the variable nature of writing traces, that impose the limitations on any inferences that can be made regarding the authorship of questioned handwriting. So where is the research headed to validate the model that is being proposed here?

### 11. Research directions

The decision as to similarity or difference is a primary candidate for research into *objective static analysis techniques* to aid in the decision process. As with the above-mentioned rationale, this research is not focused on issues of authorship, but on providing examiners with objective criteria to supplement the subjective assessment of whether a population of images is similar to or different from another population of images. Signature formations have been the initial subjects of this kind of research (Found, Metz, Rogers, Schmittat, Black & Ganas, 1994.) Signatures are straightforward to investigate in this environment because we are making inferences, mounted on its consistency and complexity, about the plausibility of a single questioned image being

the product of the same neuromuscular processes as was used to form the standard images. We can, therefore, construct at least spatial criteria that have to be met in order to proceed to the stage of proposing hypotheses about the explanations as to why an image is similar or different. An example of this kind of approach would be that in order to express the opinion that the questioned image was written by the standard writer, the signature would be required to reach a spatial criteria consistent with the population of standard images and fulfill other criteria such as those associated with complexity and subjective line quality assessments. Although this approach is theoretically and practically achievable, the research is still in its infancy. There are problems, however, in translating research on common images to examinations of extended text. The limiting factors are that we observe a phenomena thought similar to *context specific variation* for speech (discussed in Schmidt, 1988, p.238.) That is, we observe structural variation within and between characters according to their placement within word formations and/or the surrounding characters. This, coupled with a lack of objective analysis techniques that can make the required measurements efficiently, poses a challenge for the application of measurement techniques in this area.

### 12. Conclusion

As with any opinion expressed on the outcome of human movements there is a fundamental requirement to be familiar with the normal range and variation of movement outcomes in the population from which routine examination material is drawn. For handwriting examiners, this experience comes mainly from the exposure we have to handwrite throughout the course of our life, the majority of which normally would occur before specializing in forensic handwriting examination. Forensic training serves to focus our approach to the comparison process according to the method. It should not be seen to be isolated from the real basis on which our opinions are formed which is a general exposure to the population of writing images, coupled with a knowledge of the limitations of the technique and the relationship between neural representations, artifacts of movement, complexity of images, and what can reasonably be said regarding

authorship of entries based on these elements.

Handwriting examination has traditionally been a study that has developed in relative scientific isolation. The field is small and the emphasis, as we would expect, has been on application, as this is why forensic handwriting examination came about. Research and validation have suffered as a result. It has become clear that as practitioners dealing with the artifacts of human movement, we share a great amount of common ground with scientists working in mainstream paradigms. It is unlikely, however, that forensic handwriting examination will ever be considered as a science similar to these traditional scientific paradigms. The results of the Daubert hearing, given the type of information that they were provided with, appears reasonable almost to the point of generosity. The future for our profession is based on learning from the types of criticisms that have been raised and recognizing that some of the traditional beliefs in the field must be abandoned.

Only a small number of the points raised during the Daubert hearing have been discussed here. It is not suggested that the approach outlined in this paper provides a quick fix to the problems that our field is experiencing. Indeed, what has been presented requires a great deal of work to validate in the terms that were suggested by the scientists giving evidence in the hearing. That the expertise of document examiners is properly characterized as “practical in character” rather than scientific we do not consider to be inaccurate or inappropriate. However, what is important is that in common with scientific practice we present results in a way that reflects the type of information that we deal with, and respects the limitations of the assumptions and techniques we use to reach those results. Furthermore, the future of the field will ride on the back of scientific research and the criticisms raised can only aid us in attracting suitably qualified individuals and funding to carry out the required work.

### 13. References

Baier, P.E. (1995). Image processing of forensic documents. Proceedings of the Third International Conference on document Analysis and Recognition (pp .1-4). Montreal, Canada: IEEE Computer Society Press.

- Baxendale, D., & Renshaw, I.D. (1979). The large scale searching of handwriting samples. *Journal of the Forensic Science Society*, 19, 245-251.
- Beck, J. (1995). Sources of error in forensic handwriting evaluation. *Journal of Forensic Sciences*, 40, 78-82.
- Cheung, Y.I., & Leung, S.C. (1989). A comparative approach to examination of Chinese handwriting. Part 4-Identification by statistical classification techniques. *Journal of the Forensic Science Society*, 29, 77-78.
- Conway, J.V.P. (1959). *Evidential Documents*. Illinois: Charles C. Thomas.
- Eldridge, M.A., Nimmo-Smith, I., Wing, A.M., & Totty, R.N. ( 1984). The variability of selected features in cursive handwriting: categorical measures. *Journal of the Forensic Science Society*, 24, 179-219.
- Ellen, D. (1989). *The Scientific Examination of Documents: Methods and Techniques*. West Sussex: Ellis Horwood Limited.
- Found, B., & Dick, D. (1992). The structure of handwriting comparisons. An overview of the limitations within the current methodology and a discussion of an alternative method. Paper presented at the meeting of the 11th Australian and New Zealand International Symposium of the Forensic Sciences, Hobart, Tasmania.
- Found, B., Dick, D., & Rogers, D. (1994). The structure of forensic handwriting and signature comparisons. *Forensic Linguistics. The International Journal of Speech Language and the Law*, 1, 183-196.
- Found, B., Metz, H., Rogers, D., Schmittat, R., Black, D., & Ganas, J. (1994, November). An analysis of Spatial Errors in freehand simulations. Poster presented at the meeting of the 12th Australian and New Zealand International Symposium of the Forensic Sciences, Auckland, New Zealand.
- Found, B., Rogers, D., & Schmittat, R. (1994). A computer program designed to compare the spatial elements of handwriting. *Forensic Science International*, 68, 195-203.
- Found, B., & Rogers, D. (1995). Investigation of signature complexity for forensic purposes. Proceedings of the Annual Symposium of the Association of Forensic Document Examiners and the Seventh Biennial Conference of the International Graphonomics Society (pp. 52-53). Ontario, Canada: Phylmar Associates.
- Found, B., Rogers, D., & Schmittat, R. (1995). Techniques for overcoming difficulties taking objective spatial measurements from handwriting. Proceedings of the Annual Symposium of the Association of Forensic Document Examiners and the Seventh Biennial Conference of the



## Contemporary issues –A discussion of key issues in the wake of the Starzecpyzel decision - 21

- International Graphonomics Society (pp. 54-55). Ontario, Canada: Phylmar Associates.
- Found, B., Rogers, D., Schmittat, R., & Metz, H. (1994, November). A computer technique for objectively selecting measurement points from handwriting. Paper presented at the meeting of the 12th Australian and New Zealand International Symposium of the Forensic Sciences, Auckland, New Zealand.
- Franks, J.E., Davis, T.R., Totty, R.N., Hardcastle, R.A., & Grove, D.M. (1985). Variability of stroke direction between left and right handed writers. *Journal of the Forensic Science Society*, 25, 353-370.
- Hardcastle, R.A., & Kemmenoe, D. (1990). A computer-based system for the classification of handwriting on cheques. *Journal of the Forensic Science Society*, 30, 97-103.
- Hardcastle, R.A., Thornton, D., & Totty, R.N. (1986). A computer based system for the classification of handwriting on cheques. *Journal of the Forensic Science Society*, 26, 383-392.
- Harrison, W. R (1958). *Suspect Documents, their Scientific Examination*. New York : Praeger.
- Harvey, R., & Mitchell, R.M. (1973). The Nicola Brazier Murder: The role of handwriting in a large-scale investigation. *Journal of the Forensic Science Society*, 13, 157-168.
- Hecker, M.A. (1995). Gender identification through handwriting. Proceedings of the Annual Symposium of the Association of Forensic Document Examiners and the Seventh Biennial Conference of the International Graphonomics Society (pp. 74- 75). Ontario, Canada : Phylmar Associates.
- Hilton, O. (1982). *Scientific Examination of Questioned Documents*. New York : Elsevier Science Publishing Co., Inc..
- Huber, R.A., & Headrick, A.M. (1990). Let's do it by numbers. *Forensic Science International*, 46, 209-218.
- Lacey, K., & Dick, D. (1992). Class and individual characteristics. Do they have a place in handwriting identification. Poster presented at the meeting of the 11th Australian and New Zealand International Symposium of the Forensic Sciences, Hobart, Tasmania.
- Livingstone, O.B. (1962). Frequency of certain characteristics in handwriting, pen-printing of two hundred people. *Journal of Forensic Sciences* 8, 250-259.
- McAlexander, T., (1991). The standardization of handwriting opinion terminology. *Journal of Forensic Sciences*. 36, 311- 319.
- Metz, H., Found, B., Dick, D., & Rogers, D., (1994, November). The common forensic handwriting and signature method. Paper presented at the meeting of the Australian Society of Forensic Document Examiners, Wellington, New Zealand.
- Muehlberger, R.J., Newman, K.W., Regent, J., & Wichmann, J.G. (1976). A statistical examination of selected handwriting characteristics. *Journal of Forensic Sciences*, 22, 206-215.
- Osborn, A. S. (1929). *Questioned Documents* (2nd ed.). Chicago: Nelson-Hall Co..
- Proceedings of the Annual Symposium of the Association of Forensic Document Examiners and the Seventh Biennial Conference of the International Graphonomics Society. Ontario, Canada : Phylmar Associates.
- Proceedings of the Third International Conference on Document Analysis and Recognition. Montreal, Canada : IEEE Computer Society Press.
- Risinger, D.M., Denbeaux, M.P., & Saks, M.J. (1989). Exorcism of ignorance as a proxy for rational knowledge: The lessons of handwriting identification "expertise". *University of Pennsylvania Law Review*. 137, 731-792.
- Schmidt, R.A. (1988). *Motor Control and Learning* (2nd ed.). Illinois : Human Kinetics Books.
- Thomassen, A.J.W.M., & van Galen, G.P. (1992). Handwriting as a motor task: Experimentation, modeling, and simulation. In J.J. Summers (Ed.), *Approaches to the Study of Motor Control and Learning* (pp. 113-144). North Holland : Elsevier Science Publishers B.V..
- United States v. Starzecpyzel, 880 F. Supp. 1027 (S.D.N.Y. 1995)
- van Galen, G.P. (1980). Handwriting and drawing: a two-stage model of complex motor behaviour. In G.E. Stelmach & J. Requin (Eds.), *Tutorials in motor behaviour*. Amsterdam: North Holland.
- Wing, A.M., & Nimmo-Smith, I. (1987). The variability of cursive handwriting measures defined along a continuum: Letter specificity. *Journal of the Forensic Science Society*, 27, 297- 306.

